

MAMMALIAN GENES INVOLVED IN VIRAL INFECTION
AND TUMOR SUPPRESSION

ins. ε'
ins. α'
 $\alpha' > \epsilon'$

BACKGROUND

Field of the Invention

The present invention provides methods of identifying cellular genes used for viral growth or for tumor progression. Thus, the present invention relates to nucleic acids related to and methods of reducing or preventing viral infection and for suppressing tumor progression. The invention also relates to methods for screening for additional such genes.

Background art

Various projects have been directed toward isolating and sequencing the genome of various animals, notably the human. However, most methodologies provide nucleotide sequences for which no function is linked or even suggested, thus limiting the immediate usefulness of such data.

The present invention, in contrast, provides methods of screening only for nucleic acids that are involved in a specific process, *i.e.*, viral infection or tumor progression. For viral infection, the nucleic acids isolated are useful in treatments for these processes because by this method only nucleic acids which are also nonessential to the cell are isolated. Such methods are highly useful, since they ascribe a function to each isolated gene, and thus the isolated nucleic acids can immediately be utilized in various specific methods and procedures.

For, example, the present invention provides methods of isolating nucleic acids encoding gene products used for viral infection, but nonessential to the cell. Viral infections are significant causes of human morbidity and mortality. Understanding the molecular mechanisms of such infections will lead to new approaches in their treatment and control.

Viruses can establish a variety of types of infection. These infections can be generally classified as lytic or persistent, though some lytic infections are considered persistent. Generally, persistent infections fall into two categories: (1) chronic (productive) infection, *i.e.*, infection wherein infectious virus is present and can be recovered by traditional biological methods and (2) latent infection, *i.e.*, infection

wherein viral genome is present in the cell but infectious virus is generally not produced except during intermittent episodes of reactivation. Persistence generally involves stages of both productive and latent infection.

Lytic infections can also persist under conditions where only a small fraction of the total cells are infected (smoldering (cycling) infection). The few infected cells release virus and are killed, but the progeny virus again only infect a small number of the total cells. Examples of such smoldering infections include the persistence of lactic dehydrogenase virus in mice (Mahy, B.W.J., *Br. Med. Bull.* 41: 50-55 (1985)) and adenovirus infection in humans (Porter, D.D. pp. 784-790 in Baron, S., ed. *Medical Microbiology* 2d ed. (Addison-Wesley, Menlo Park, CA 1985)).

Furthermore, a virus may be lytic for some cell types but not for others. For example, evidence suggests that human immunodeficiency virus (HIV) is more lytic for T cells than for monocytes/macrophages, and therefore can result in a productive infection of T cells that can result in cell death, whereas HIV-infected mononuclear phagocytes may produce virus for considerable periods of time without cell lysis. (Klatzmann, et al. *Science* 225:59-62 (1984); Koyanagi, et al. *Science* 241:1673-1675 (1988); Sattentau, et al. *Cell* 52:631-633 (1988)).

Traditional treatments for viral infection include pharmaceuticals aimed at specific virus derived proteins, such as HIV protease or reverse transcriptase, or recombinant (cloned) immune modulators (host derived), such as the interferons. However, the current methods have several limitations and drawbacks which include high rates of viral mutations which render anti-viral pharmaceuticals ineffective. For immune modulators, limited effectiveness, limiting side effects, a lack of specificity all limit the general applicability of these agents. Also the rate of success with current antivirals and immune-modulators has been disappointing.

One aspect of the current invention focuses on isolating genes that are not essential for cellular survival when disrupted in one or both alleles, but which are required for virus replication. This may occur with a dose effect, in which one allele knock-out may confer the phenotype of virus resistance for the cell. As targets for therapeutic intervention, inhibition of these cellular gene products, including: proteins, parts of proteins (modification enzymes that include, but are not restricted to glycosylation, lipid modifiers [myriolate, etc.]), lipids, transcription elements and RNA

regulatory molecules, may be less likely to have profound toxic side effects and virus mutation is less likely to overcome the 'block' to replicate successfully.

The present invention provides a significant improvement over previous methods of attempted therapeutic intervention against viral infection by addressing the cellular
5 genes required by the virus for growth. Therefore, the present invention also provides an innovative therapeutic approach to intervention in viral infection by providing methods to treat viruses by inhibiting the cellular genes necessary for viral infection. Because these genes, by virtue of the means by which they are originally detected, are nonessential to the cell's survival at a level of expression necessary to inhibit virus
10 replication, these treatment methods can be used in a subject without serious detrimental effects to the subject, as has been found with previous methods. The present invention also provides the surprising discovery that virally infected cells are dependent upon a factor in serum to survive. Therefore, the present invention also provides a method for treating viral infection by inhibiting this serum survival factor. Finally, these discoveries
15 also provide a novel method for removing virally infected cells from a cell culture by removing, inhibiting or disrupting this serum survival factor in the culture so that non-infected cells selectively survive.

The selection of tumor suppressor gene(s) has become an important area in the discovery of new target for therapeutic intervention of cancer. Since the discovery that
20 cells are restricted from promiscuous entry into the cell cycle by specific genes that are capable of suppressing a 'transformed' phenotype, considerable time has been invested in the discovery of such genes. Some of these genes include the gene associated by rhabdomyosarcoma (Rb) and the p53 (apoptosis related) encoding gene. The present invention provides a method, using gene-trapping, to select cell lines that have a
25 transformed phenotype from cells that are not transformed and to isolate from these cells a gene that can suppress a malignant, or transformed, phenotype. Thus, by the nature of the isolation process, a function is associated with the isolated genes. The capacity to select quickly tumor suppressor genes can provide unique targets in the process of treating or preventing, and even for diagnostic testing of, cancer.

DETAILED DESCRIPTION OF THE INVENTION

The present invention utilizes a "gene trap" method along with a selection process to identify and isolate nucleic acids from genes associated with a particular function. Specifically, it provides a means of isolating cellular genes necessary for viral infection but not essential for the cell's survival, and it provides a means of isolating cellular genes that suppress tumor progression.

The present invention also provides a core discovery that virally infected cells become dependent upon at least one factor present in serum for survival, whereas non-infected cells do not exhibit this dependence. This core discovery has been utilized in the present invention in several ways. First, inhibition of the "serum survival factor" can be utilized to eradicate persistently virally infected cells from populations of non-infected cells. Inhibition of this factor can also be used to treat virus infection in a subject, as further described herein. Additionally, inhibition of or withdrawal of the serum survival factor in tissue culture allows for the detection of cellular genes required for viral replication yet nonessential for an uninfected cell to survive. The present invention further provides several such cellular genes, as well as methods of treating viral infections by inhibiting the functioning of such genes.

The invention also provides cellular genes whose overexpression is associated with inhibition of viral growth and/or reproduction.

The present method provides several cellular genes that are necessary for viral growth in the cell but are not essential for the cell to survive. These genes are important for lytic and persistent infection by viruses. These genes were isolated by generating gene trap libraries by infecting cells with a retrovirus gene trap vector, selecting for cells in which a gene trap event occurred (*i.e.*, in which the vector had inserted such that the promoterless marker gene was inserted such that a cellular promoter promotes transcription of the marker gene, *i.e.*, inserted into a functioning gene), starving the cells of serum, infecting the selected cells with the virus of choice while continuing serum starvation, and adding back serum to allow visible colonies to develop, which colonies were cloned by limiting dilution. Genes into which the retrovirus gene trap vector inserted were then isolated from the colonies using probes specific for the retrovirus gene trap vector. Thus nucleic acids isolated by this method are isolated portions of genes. Additionally, utilizing this method, several cellular genes were isolated whose

overexpression prevents viral infection or tumor growth, and they provide methods of treating viral infection or tumor growth/suppression by overexpression of these genes.

Thus the present invention provides a method of identifying a cellular gene necessary for viral growth in a cell and nonessential for cellular survival, comprising (a) transferring into a cell culture, *e.g.* growing in serum-containing medium, a vector encoding a selective marker gene lacking a functional promoter, (b) selecting cells expressing the marker gene, (c) removing serum from the culture medium, (d) infecting the cell culture with the virus, and (e) isolating from the surviving cells a cellular gene within which the marker gene is inserted, thereby identifying a gene necessary for viral growth in a cell and nonessential for cellular survival. The present invention also provides a method of identifying a cellular gene used for viral growth in a cell and nonessential for cellular survival, comprising (a) transferring into a cell culture growing in serum-containing medium a vector encoding a selective marker gene lacking a functional promoter, (b) selecting cells expressing the marker gene, (c) removing serum from the culture medium, (d) infecting the cell culture with the virus, and (e) isolating from the surviving cells a cellular gene within which the marker gene is inserted, thereby identifying a gene necessary for viral growth in a cell and nonessential for cellular survival or a gene whose overexpression prevents viral reproduction but is not fatal to the survival to the cell. In any selected cell type, such as Chinese hamster ovary cells, one can readily determine if serum starvation is required for selection. If it is not, serum starvation may be eliminated from the steps.

Alternatively, instead of removing serum from the culture medium, a serum factor required by the virus for growth can be inhibited, such as by the administration of an antibody that specifically binds that factor. Furthermore, if it is believed that there are no persistently infected cells in the culture, the serum starvation step can be eliminated and the cells grown in usual medium for the cell type. If serum starvation is used, it can be continued for a time after the culture is infected with the virus. Serum can then be added back to the culture. If some other method is used to inactivate the factor, it can be discontinued, inactivated or removed (such as removing the anti-factor antibody, *e.g.*, with a bound antibody directed against that antibody) prior to adding fresh serum back to the culture. Cells that survive are mutants having an inactivating insertion in a gene necessary for growth of the virus. The genes having the insertions can then be isolated by isolating sequences having the marker gene sequences. This mutational process

disturbs a wild type function. A mutant gene may produce at a lower level a normal product, it may produce a normal product not normally found in these cells, it may cause the overproduction of a normal product, it may produce an altered product that has some functions but not others, or it may completely disrupt a gene function. Additionally, the mutation may disrupt an RNA that has a function but is never translated into a protein. For example, the alpha-tropomyosin gene has a 3' RNA that is very important in cell regulation but never is translated into protein. (*Cell* 75 pg 1107-1117, 12/17/93).

As used herein, a cellular gene "nonessential for cellular survival" means a gene for which disruption of one or both alleles results in a cell viable for at least a period of time which allows viral replication to be inhibited for preventative or therapeutic uses or use in research. A gene "necessary for viral growth" means the gene product, either protein or RNA, secreted or not, is necessary or beneficial, either directly or indirectly in some way for the virus to grow, and therefore, in the absence of that gene product (*i.e.*, a functionally available gene product), the virus does not spread. For example, such genes can encode cell cycle regulatory proteins, proteins affecting the vacuolar hydrogen pump, or proteins involved in protein folding and protein modification, including but not limited to: phosphorylation, methylation, glycosylation, myristylation or other lipid moiety, or protein processing via enzymatic processing. Some examples of such genes include vacuolar H⁺ATPase, alpha tropomyosin, gas5 gene, ras complex, N-acetylglucosaminyl-transferase I mRNA, annexin II, c-golgi CM130 and calcyclin.

Any virus capable of infecting the cell can be used for this method. Virus can be selected based upon the particular infection desired to study. However, it is contemplated by the present invention that many viruses will be dependent upon the same cellular genes for survival; thus a cellular gene isolated using one virus can be used as a target for therapy for other viruses as well. Any cellular gene can be tested for relevancy to any desired virus using the methods set forth herein, *i.e.*, in general, by inhibiting the gene or its gene product in a cell and determining if the desired virus can grow in that cell. Some examples of viruses include HIV (including HIV-1 and HIV-2); parvovirus; papillomaviruses; hantaviruses; influenza viruses (*e.g.*, influenza A, B and C viruses); hepatitis viruses A to G; caliciviruses; astroviruses; rotaviruses; coronaviruses, such as human respiratory coronavirus; picornaviruses, such as human rhinovirus and enterovirus; ebola virus; human herpesvirus (*e.g.*, HSV-1-9); human adenovirus; for animal, the animal counterpart to any above listed human virus, animal retroviruses, such

as simian immunodeficiency virus, avian immunodeficiency virus, bovine immunodeficiency virus, feline immunodeficiency virus, equine infectious anemia virus, caprine arthritis encephalitis virus, arenaviruses, arvoviruses, tickborne viruses or visna virus.

5 The nucleic acids comprising cellular genes of this invention were isolated by the above method and as set forth in the examples. The invention includes a nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, 10 SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, 15 SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, 20 SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:91, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:112, SEQ 25 ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, and SEQ ID NO:127 (this list is sometimes referred to herein as "SEQ LIST 1" for brevity). Thus these nucleic acids can contain, in addition to the nucleotides set forth in each SEQ ID NO in the sequence listing, additional nucleotides at either end of the molecule. Such additional nucleotides can be added by 30 any standard method, as known in the art, such as recombinant methods and synthesis methods. Examples of such nucleic acids comprising the nucleotide sequence set forth in any entry of the sequence listing contemplated by this invention include, but are not limited to, for example, the nucleic acid placed into a vector; a nucleic acid having one or

more regulatory region (*e.g.*, promoter, enhancer, polyadenylation site) linked to it, particularly in functional manner, *i.e.* such that an mRNA or a protein can be produced; a nucleic acid including additional nucleic acids of the gene, such as a larger or even full length genomic fragment of the gene, a partial or full length cDNA, a partial or full length RNA. Making and/or isolating such larger nucleic acids is further described below and is well known and standard in the art.

Also provided in this invention are the double-stranded nucleic acids corresponding to the nucleic acid sequences set forth in SEQ ID 1 through SEQ ID 136, inclusive. It is recognized that "nucleic acid" as used herein, can refer to either or both strands of such double-stranded nucleic acids, such strands often referred to as the "positive" and "negative" strands. Either strand of such double-stranded nucleic acids may encode the polypeptides of this invention, and the coding sequences for such polypeptides may be translated in either direction along the strand. Examples of polypeptides encoded by either strand are disclosed herein.

The invention also provides a nucleic acid encoding the protein encoded by the gene comprising the nucleotide sequence set forth in any of the sequences listed in SEQ LIST 1, as well as allelic variants and homologs of each such gene. The gene is readily obtained using standard methods, as described below and as is known and standard in the art. The present invention also contemplates any unique fragment of these genes or of the nucleic acids set forth in any of the sequences listed in SEQ LIST 1. Examples of inventive fragments of the inventive genes can include the nucleic acids whose sequence is set forth in any of the sequences listed in SEQ LIST 1. To be unique, the fragment must be of sufficient size to distinguish it from other known sequences, most readily determined by comparing any nucleic acid fragment to the nucleotide sequences of nucleic acids in computer databases, such as GenBank. Such comparative searches are standard in the art. Typically, a unique fragment useful as a primer or probe will be at least about 20 to about 25 nucleotides in length, depending upon the specific nucleotide content of the sequence. Additionally, fragments can be, for example, at least about 30, 40, 50, 75, 100, 200 or 500 nucleotides in length. The nucleic acids can be single or double stranded, depending upon the purpose for which it is intended.

The present invention further provides a nucleic acid comprising the regulatory region of a gene comprising any one of the nucleotide sequences set forth in SEQ LIST 1, as well as homologs of each such gene. Additionally provided is a construct

comprising such a regulatory region functionally linked to a reporter gene. Such reporter gene constructs can be used to screen for compounds and compositions that affect expression of the gene comprising the nucleic acids whose sequence is set forth in SEQ LIST 1, or any homologs thereof.

- 5 The nucleic acids set forth in the sequence listing are gene fragments; the entire coding sequence and the entire gene that comprises each fragment are both contemplated herein and are readily obtained by standard methods, given the nucleotide sequences presented in the sequence listing (*see. e.g.*, Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989; *DNA cloning: A Practical Approach*, Volumes I and II, Glover, D.M. ed., IRL Press Limited, Oxford, 1985). To obtain the entire genomic gene, briefly, a nucleic acid whose sequence is set forth in any of SEQ ID NO:1 through SEQ ID NO:127, or preferably in any of the sequences listed in SEQ LIST 1, or a smaller fragment thereof, is utilized as a probe to screen a genomic library under high stringency conditions, and
- 10 isolated clones are sequenced. Once the sequence of the new clone is determined, a probe can be devised from a portion of the new clone not present in the previous fragment and hybridized to the library to isolate more clones containing fragments of the gene. In this manner, by repeating this process in organized fashion, one can "walk" along the chromosome and eventually obtain nucleotide sequence for the entire gene.
- 15 Similarly, one can use portions of the present fragments, or additional fragments obtained from the genomic library, that contain open reading frames to screen a cDNA library to obtain a cDNA having the entire coding sequence of the gene. Repeated screens can be utilized as described above to obtain the complete sequence from several clones if necessary. The isolates can then be sequenced to determine the nucleotide
- 20 sequence by standard means such as dideoxynucleotide sequencing methods (*see, e.g.*, Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989).

- The present genes were isolated from rat; however, homologs in any desired species, preferably mammalian, such as human, can readily be obtained by screening a
- 30 human library, genomic or cDNA, with a probe comprising sequences of the nucleic acids set forth in the sequence listing herein, or fragments thereof, and isolating genes specifically hybridizing with the probe under preferably relatively high stringency hybridization conditions. For example, high salt conditions (*e.g.*, in 6X SSC or 6X

SSPE) and/or high temperatures of hybridization can be used. For example, the stringency of hybridization is typically about 5°C to 20°C below the T_m (the melting temperature at which half of the molecules dissociate from its partner) for the given chain length. As is known in the art, the nucleotide composition of the hybridizing region factors in determining the melting temperature of the hybrid. For 20mer probes, for example, the recommended hybridization temperature is typically about 55-58°C. Additionally, the rat sequence can be utilized to devise a probe for a homolog in any specific animal by determining the amino acid sequence for a portion of the rat protein, and selecting a probe with optimized codon usage to encode the amino acid sequence of the homolog in that particular animal. Any isolated gene can be confirmed as the targeted gene by sequencing the gene to determine it contains the nucleotide sequence listed herein as comprising the gene. Any homolog can be confirmed as a homolog by its functionality.

Additionally contemplated by the present invention are nucleic acids, from any desired species, preferably mammalian and more preferably human, having 98%, 95%, 90%, 85%, 80%, 70%, 60%, or 50% homology, or greater, in the region of homology, to a region in an exon of a nucleic acid encoding the protein encoded by the gene comprising the nucleotide sequence set forth in any of the sequences listed in SEQ LIST 1 or to homologs thereof. Also contemplated by the present invention are nucleic acids, from any desired species, preferably mammalian and more preferably human, having 98%, 95%, 90%, 85%, 80%, 70%, 60%, or 50% homology, or greater, in the region of homology, to a region in an exon of a nucleic acid comprising the nucleotide sequence set forth in any of the sequences listed in SEQ LIST 1 or to homologs thereof. These genes can be synthesized or obtained by the same methods used to isolate homologs, with stringency of hybridization and washing, if desired, reduced accordingly as homology desired is decreased, and further, depending upon the G-C or A-T richness of any area wherein variability is searched for. Allelic variants of any of the present genes or of their homologs can readily be isolated and sequenced by screening additional libraries following the protocol above. Methods of making synthetic genes are described in U.S. Patent No. 5,503,995 and the references cited therein.

The nucleic acid encoding any selected protein of the present invention can be any nucleic acid that functionally encodes that protein. For example, to functionally encode, *i.e.*, allow the nucleic acid to be expressed, the nucleic acid can include, for

example, exogenous or endogenous expression control sequences, such as an origin of replication, a promoter, an enhancer, and necessary information processing sites, such as ribosome binding sites, RNA splice sites, polyadenylation sites, and transcriptional terminator sequences. Preferred expression control sequences can be promoters derived
5 from metallothioneine genes, actin genes, immunoglobulin genes, CMV, SV40, adenovirus, bovine papilloma virus, etc. Expression control sequences can be selected for functionality in the cells in which the nucleic acid will be placed. A nucleic acid encoding a selected protein can readily be determined based upon the amino acid sequence of the selected protein, and, clearly, many nucleic acids will encode any selected
10 protein.

The present invention additionally provides a nucleic acid that selectively hybridizes under stringent conditions with a nucleic acid set forth in SEQ LIST 1 or with a nucleic acid encoding the protein encoded by the gene comprising the nucleotide sequence set forth in any sequence listed in SEQ LIST 1. This hybridization can be
15 specific. The degree of complementarity between the hybridizing nucleic acid and the sequence to which it hybridizes should be at least enough to exclude hybridization with a nucleic acid encoding an unrelated protein. Thus, a nucleic acid that selectively hybridizes with a nucleic acid of the present protein coding sequence will not selectively hybridize under stringent conditions with a nucleic acid for a different, unrelated protein,
20 and vice versa. Typically, the stringency of hybridization to achieve selective hybridization involves hybridization in high ionic strength solution (6X SSC or 6X SSPE) at a temperature that is about 12-25°C below the T_m (the melting temperature at which half of the molecules dissociate from its partner) followed by washing at a combination of temperature and salt concentration chosen so that the washing temperature is about 5°C
25 to 20°C below the T_m of the hybrid molecule. The temperature and salt conditions are readily determined empirically in preliminary experiments in which samples of reference DNA immobilized on filters are hybridized to a labeled nucleic acid of interest and then washed under conditions of different stringencies. Hybridization temperatures are typically higher for DNA-RNA and RNA-RNA hybridizations. The washing
30 temperatures can be used as described above to achieve selective stringency, as is known in the art. (Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989; Kunkel et al. *Methods Enzymol.* 1987:154:367, 1987). Nucleic acid fragments that selectively hybridize to any

given nucleic acid can be used, *e.g.*, as primers and or probes for further hybridization or for amplification methods (*e.g.*, polymerase chain reaction (PCR), ligase chain reaction (LCR)). A preferable stringent hybridization condition for a DNA:DNA hybridization can be at about 68°C (in aqueous solution) in 6X SSC or 6X SSPE followed by washing
5 at 68°C.

The present invention additionally provides a polypeptide comprising the amino acid sequence encoded by the gene comprising the nucleotide sequence set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ
10 ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID
15 NO:45, SEQ ID NO:46, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID
20 NO:80, SEQ ID NO:81, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:91, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108,
25 SEQ ID NO:112, SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, and SEQ ID NO:127 (*i.e.*, SEQ LIST 1). Additionally, polypeptides comprising the amino acid sequence encoded by a nucleic acid that selectively hybridizes under stringent conditions with a nucleic acid in SEQ LIST 1 are provided. Further, polypeptides comprising the amino acid sequence
30 encoded by a nucleic acid having a region within an exon wherein the region has at least 50, 60, 70, 80, 90, or 95% homology with a nucleic acid in SEQ LIST 1. These polypeptides can be readily obtained by any of several means. For example, the nucleotide sequence of coding regions of the gene can be translated and then the

corresponding polypeptide can be synthesized mechanically by standard methods. Additionally, the coding regions of the genes can be expressed or synthesized, an antibody specific for the resulting polypeptide can be raised by standard methods (see, e.g., Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1988), and the protein can be isolated from other cellular proteins by selective hybridization with the antibody. This protein can be purified to the extent desired by standard methods of protein purification (see, e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989). The amino acid sequence of any protein, polypeptide or peptide of this invention can be deduced from the nucleic acid sequence, or it can be determined by sequencing an isolated or recombinantly produced protein.

The terms "peptide," "polypeptide" and "protein" can be used interchangeably herein and refer to a polymer of amino acids and includes full-length proteins and fragments thereof. As used in the specification and in the claims, "a" can mean one or more, depending upon the context in which it is used. An amino acid residue is an amino acid formed upon chemical digestion (hydrolysis) of a polypeptide at its peptide linkages. The amino acid residues described herein are preferably in the L isomeric form. However, residues in the D isomeric form can be substituted for any L-amino acid residue, as long as the desired functional property is retained by the polypeptide. Standard polypeptide nomenclature (described in *J. Biol. Chem.*, 243:3552-59 (1969) and adopted at 37 CFR § 1.822(b)) is used herein.

As will be appreciated by those skilled in the art, the invention also includes those polypeptides having slight variations in amino acid sequences or other properties. Amino acid substitutions can be selected by known parameters to be neutral (see, e.g., Robinson WE Jr, and Mitchell WM., *AIDS* 4:S151-S162(1990)). Such variations may arise naturally as allelic variations (e.g., due to genetic polymorphism) or may be produced by human intervention (e.g., by mutagenesis of cloned DNA sequences), such as induced point, deletion, insertion and substitution mutants. Minor changes in amino acid sequence are generally preferred, such as conservative amino acid replacements, small internal deletions or insertions, and additions or deletions at the ends of the molecules. Substitutions may be designed based on, for example, the model of Dayhoff, *et al.* (in *Atlas of Protein Sequence and Structure* 1978, Nat'l Biomed. Res. Found., Washington,

D.C.). These modifications can result in changes in the amino acid sequence, provide silent mutations, modify a restriction site, or provide other specific mutations. Likewise, such amino acid changes result in a different nucleic acid encoding the polypeptides and proteins. Thus, alternative nucleic acids are also contemplated by such modifications.

5 The present invention also provides cells containing a nucleic acid of the invention. A cell containing a nucleic acid encoding a protein typically can replicate the DNA and, further, typically can express the encoded protein. The cell can be a prokaryotic cell, particularly for the purpose of producing quantities of the nucleic acid, or a eukaryotic cell, particularly a mammalian cell. The cell is preferably a mammalian
10 cell for the purpose of expressing the encoded protein so that the resultant produced protein has mammalian protein processing modifications.

Nucleic acids of the present invention can be delivered into cells by any selected means, in particular depending upon the purpose of the delivery of the compound and the target cells. Many delivery means are well-known in the art. For example,
15 electroporation, calcium phosphate precipitation, microinjection, cationic or anionic liposomes, and liposomes in combination with a nuclear localization signal peptide for delivery to the nucleus can be utilized, as is known in the art.

The present invention also contemplates that the mutated cellular genes necessary for viral growth, produced by the present method, as well as cells containing these
20 mutants can also be useful. These mutated genes and cells containing them can be isolated and/or produced according to the methods herein described and using standard methods.

It should be recognized that the sequences set forth herein may contain minor sequencing errors. Such errors can be corrected, for example, by using the hybridization
25 procedure described above with various probes derived from the described sequences such that the coding sequence can be reisolated and resequenced.

As described in the examples, the present invention provides the discovery of a "serum survival factor" present in serum that is necessary for the survival of persistently virally infected cells. Isolation and characterization of this factor have shown it to be a
30 protein, to have a molecular weight of between about 50 kD and 100 kD, to resist inactivation in low pH (*e.g.*, pH2) and chloroform extraction, to be inactivated by boiling for about 5 minutes and in low ionic strength solution (*e.g.*, about 10 mM to about 50 mM). The present invention thus provides a purified mammalian serum protein having a

molecular weight of between about 50 kD and 100 kD which resists inactivation in low pH and resists inactivation by chloroform extraction, which inactivates when boiled and inactivates in low ionic strength solution, and which when removed from a cell culture comprising cells persistently infected with reovirus selectively substantially prevents survival of cells persistently infected with reovirus. The factor, fitting the physical characteristics described above, can readily be verified by adding it to non-serum-containing medium (which previously could not support survival of persistently virally infected cells) and determining whether this medium with the added putative factor can now support persistently virally infected cells, particularly cells persistently infected with reovirus. As used herein, a "purified" protein means the protein is at least of sufficient purity such that an approximate molecular weight can be determined.

The amino acid sequence of the protein can be elucidated by standard methods. For example, an antibody to the protein can be raised and used to screen an expression library to obtain nucleic acid sequence coding the protein. This nucleic acid sequence is then simply translated into the corresponding amino acid sequence. Alternatively, a portion of the protein can be directly sequenced by standard amino acid sequencing methods (amino-terminus sequencing). This amino acid sequence can then be used to generate an array of nucleic acid probes that encompasses all possible coding sequences for a portion of the amino acid sequence. The array of probes is used to screen a cDNA library to obtain the remainder of the coding sequence and thus ultimately the corresponding amino acid sequence.

The present invention also provides methods of detecting and isolating additional serum survival factors. For example, to determine if any known serum components are necessary for viral growth, the known components can be inhibited in, or eliminated from, the culture medium, and it can be observed whether viral growth is inhibited by determining if persistently infected cells do not survive. One can add the factor back (or remove the inhibition) and determine whether the factor allows for viral growth.

Additionally, other, unknown serum components can also be found to be essential for growth. Serum can be fractionated by various standard means, and fractions added to serum free medium to determine if a factor is present in a reaction that allows growth previously inhibited by the lack of serum. Fractions having this activity can then be further fractionated until the factor is relatively free of other components. The factor can then be characterized by standard methods, such as size fractionation, denaturation and/or

inactivation by various means, etc. Preferably, once the factor has been purified to a desired level of purity, it is added to cells in serum free medium to confirm that it bestows the function of allowing virus to grow when serum-free medium alone did not. This method can be repeated to confirm the requirement for the specific factor for any desired virus, since each serum factor found to be required by any one virus can also be required by many other viruses. In general, the closer the viruses are related and the more similar the infection modes of the viruses, the more likely that a factor required by one virus will be required by the other.

The present invention also provides methods of treating virus infections utilizing applicants' discoveries. The subject of any of the herein described methods can be any animal, preferably a mammal, such as a human, a veterinary animal, such as a cat, dog, horse, pig, goat, sheep, or cow, or a laboratory animal, such as a mouse, rat, rabbit, or guinea pig, depending upon the virus.

The present invention provides a method of reducing or inhibiting, and thereby treating, a viral infection in a subject, comprising administering to the subject an inhibiting amount of a composition that inhibits functioning of the serum protein described herein, *i.e.* the serum protein having a molecular weight of between about 50 kD and 100 kD which resists inactivation in low pH and resists inactivation by chloroform extraction, which inactivates when boiled and inactivates in low ionic strength solution, and which when removed from a cell culture comprising cells persistently infected with the virus prevents survival of at least some cells persistently infected with the virus, thereby treating the viral infection. The composition can comprise, for example, an antibody that specifically binds the serum protein, or an antisense RNA that binds an RNA encoded by a gene functionally encoding the serum protein.

Any virus capable of infecting the selected subject to be treated can be treated by the present methods. As described above, any serum protein or survival factor found by the present methods to be necessary for growth of cells infected with any one virus can be found to be necessary for growth of the cells infected with many other viruses. For any given cell-virus combination, the serum protein or factor can be confirmed to be required for growth by the methods described herein. The cellular genes identified by the examples using reovirus, a mammalian pathogen, and a rat cell system have general applicability to other virus infections that include all of the known as well as yet to be discovered human pathogens, including, but not limited to: human immunodeficiency

viruses (e.g., HIV-1, HIV-2); parvovirus; papillomaviruses; hantaviruses; influenza viruses (e.g., influenza A, B and C viruses); hepatitis viruses A to G; caliciviruses; astroviruses; rotaviruses; coronaviruses, such as human respiratory coronavirus; picornaviruses, such as human rhinovirus and enterovirus; ebola virus; human
5 herpesvirus (e.g., HSV-1-9); human adenovirus; hantaviruses; for animal, the animal counterpart to any above listed human virus, animal retroviruses, such as simian immunodeficiency virus, avian immunodeficiency virus, bovine immunodeficiency virus, feline immunodeficiency virus, equine infectious anemia virus, caprine arthritis encephalitis virus, arenaviruses, arvoviruses, tickborne virus or visna virus.

10 A protein inhibiting amount of the composition can be readily determined, such as by administering varying amounts to cells or to a subject and then adjusting the effective amount for inhibiting the protein according to the volume of blood or weight of the subject. Compositions that bind to the protein can be readily determined by running the putatively bound protein on a protein gel and observing an alteration in the protein's
15 migration through the gel. Inhibition of the protein can be determined by any desired means such as adding the inhibitor to complete media used to maintain persistently infected cells and observing the cells' viability. The composition can comprise, for example, an antibody that specifically binds the serum protein. Specific binding by an antibody means that the antibody can be used to selectively remove the factor from serum
20 or inhibit the factor's biological activity and can readily be determined by radio immune assay (RIA), bioassay, or enzyme-linked immunosorbant (ELISA) technology. The composition can comprise, for example, an antisense RNA that specifically binds an RNA encoded by the gene encoding the serum protein. Antisense RNAs can be synthesized and used by standard methods (e.g., *Antisense RNA and DNA*, D. A. Melton, Ed., Cold
25 Spring Harbor Laboratory, Cold Spring Harbor, NY (1988)).

The present methods provide a method of screening a compound for effectiveness in treating or preventing a viral infection, comprising administering the compound to a cell containing a cellular gene functionally encoding a gene product necessary for reproduction of the virus in the cell but not necessary for survival of the cell and detecting the level
30 and/or activity (i.e. function) of the gene product produced, a decrease or elimination of the gene product and/or the gene product activity indicating a compound for treating or preventing the viral infection. The cellular gene can be, for example, a nucleic acid set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ

ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82; SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117, SEQ ID NO:118, SEQ ID NO:119, SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, or SEQ ID NO:127 (herein sometimes referred to as SEQ LIST 2, for brevity), any homolog thereof, or any other gene obtained using the methods provided herein for obtaining such genes. It is understood that the cellular gene can be present naturally in the cell being screened, or it can be introduced into the cell in a suitable expression vector, as are well known in the art. The level of the gene product can be measured by any standard means, such as by detection with an antibody specific for the protein. The level of gene product can be compared to the level of the gene product in a control cell not contacted with the compound. The level of gene product can be compared to the level of the gene product in the same cell prior to addition of the compound. Activity, or function, can be measured by any standard means, such as by enzymatic assays that measure the conversion of a substrate to a product or binding

assays that measure the binding of a protein to a nucleic acid, for example. Examples of gene products disclosed herein whose activity/function can be measured include tristetraprolin (human ZFP-36), 6-pyruvoyl-tetrahydropterin synthase, a eukaryotic DnaJ-like protein, ID3 (inhibitor of DNA binding 3), N-acetylglucos-aminyltransferase I (mGAT-1), cleavage stimulation factor (CSTF2), TAK1 binding protein, human zinc transcription factor ZPF207, Dlx2, Smad7 (Mad-related protein), and P-glycoprotein (mdr1b). The activity can be compared to the activity in a control cell not contacted with the compound or in the same cell prior to addition of the compound. Relatedly, the regulatory region of the gene can be functionally linked to a reporter gene and compounds can be screened for inhibition of the reporter gene. Such reporter constructs are described herein.

The present invention also provides a method of screening a compound for effectiveness in treating or preventing a viral infection comprising contacting the compound with the gene product of a cellular gene comprising a nucleic acid of SEQ LIST 2, or any homolog thereof, and detecting the function of the gene product, a decrease or elimination of the function indicating a compound effective for treating or preventing viral infection. Examples of gene products disclosed herein that can be utilized in this method include tristetraprolin (human ZFP-36), 6-pyruvoyl-tetrahydropterin synthase, a eukaryotic DnaJ-like protein, ID3 (inhibitor of DNA binding 3), N-acetylglucos-aminyltransferase I (mGAT-1), cleavage stimulation factor (CSTF2), TAK1 binding protein, human zinc transcription factor ZPF207, Dlx2, Smad7 (Mad-related protein), and P-glycoprotein (mdr1b).

The present invention provides a method of selectively eliminating cells persistently infected with a virus from an animal cell culture capable of surviving for a first period of time in the absence of serum, comprising propagating the cell culture in the absence of serum for a second time period during which a persistently infected cell cannot survive without serum, thereby selectively eliminating from the cell culture cells persistently infected with the virus. The second time period should be shorter than the first time period. Thus one can simply eliminate serum from a standard culture medium composition for a period of time (*e.g.* by removing serum containing medium from the culture container, rinsing the cells, and adding serum-free medium back to the container), then, after a time of serum starvation, return serum to the culture medium. Alternatively, one can inhibit a serum survival factor from the culture in place of the step of serum starvation. Furthermore, one can instead interfere with the virus-factor interaction. Such a

viral elimination method can periodically be performed for cultured cells to ensure that they remain virus-free. The time period of serum removal can greatly vary, with a typical range being about 1 to about 30 days; a preferable period can be about 3 to about 10 days, and a more preferable period can be about 5 days to about 7 days. This time period can be
5 selected based upon ability of a specific cell to survive without serum as well as the life cycle of the target virus, *e.g.*, for reovirus, which has a life cycle of about 24 hours, 3 days' starvation of cells provides dramatic results.

Furthermore, the time period can be shortened by also passaging the cells during the starvation; in general, increasing the number of passages can decrease the time of
10 serum starvation (or serum factor inhibition) needed to get full clearance of the virus from the culture. While passaging, the cells typically are exposed briefly to serum (typically for about 3 to about 24 hours). This exposure both stops the action of the trypsin used to dislodge the cells and stimulates the cells into another cycle of growth, thus aiding in this selection process. Thus a starvation/serum cycle can be repeated to optimize the selective
15 effect. Other standard culture parameters, such as confluency of the cultures, pH, temperature, etc. can be varied to alter the needed time period of serum starvation (or serum survival factor inhibition). This time period can readily be determined for any given viral infection by simply removing the serum for various periods of time, then testing the cultures for the presence of the infected cells (*e.g.*, by ability to survive in the
20 absence of serum and confirmed by quantitating virus in cells by standard virus titration and immunohistochemical techniques) at each tested time period, and then detecting at which time periods of serum deprivation the virally infected cells were eliminated. It is preferable that shorter time periods of serum deprivation that still provide elimination of the persistently infected cells be used. Furthermore, the cycle of starvation, then adding
25 back serum and determining amount of virus remaining in the culture can be repeated until no virtually infected cells remain in the culture.

Thus, the present method can further comprise passaging the cells, *i.e.*, transferring the cell culture from a first container to a second container. Such transfer can facilitate the selective lack of survival of virally infected cells. Transfer can be repeated several times.
30 Transfer is achieved by standard methods of tissue culture (*see, e.g.*, Freshney, *Culture of Animal Cells, A Manual of Basic Technique*, 2nd Ed. Alan R. Liss, Inc., New York, 1987).

The present method further provides a method of selectively eliminating from a cell culture cells persistently infected with a virus, comprising propagating the cell culture

in the absence of a functional form of the serum protein having a molecular weight of between about 50 kD and 100 kD which resists inactivation in low pH and resists inactivation by chloroform extraction, which inactivates when boiled and inactivates in low ionic strength solution, and which when removed from a cell culture comprising cells persistently infected with reovirus substantially prevents survival of cells persistently infected with reovirus. The absence of the functional form can be achieved by any of several standard means, such as by binding the protein to an antibody selective for it (binding the antibody in serum either before or after the serum is added to the cells; if before, the serum protein can be removed from the serum by, *e.g.*, binding the antibody to a column and passing the serum over the column and then administering the survival protein-free serum to the cells), by administering a compound that inactivates the protein, or by administering a compound that interferes with the interaction between the virus and the protein.

Thus, the present invention provides a method of selectively eliminating from a cell culture propagated in serum-containing medium cells persistently infected with a virus, comprising inhibiting in the serum the protein having a molecular weight of between about 50 kD and 100 kD which resists inactivation in low pH and resists inactivation by chloroform extraction, which inactivates when boiled and inactivates in low ionic strength solution, and which when removed from a cell culture comprising cells persistently infected with reovirus substantially prevents survival of cells persistently infected with reovirus. Alternatively, the interaction between the virus and the serum protein can be disrupted to selectively eliminate cells persistently infected with the virus.

Any virus capable of some form of persistent infection may be eliminated from a cell culture utilizing the present elimination methods, including removing, inhibiting or otherwise interfering with a serum protein, such as the one exemplified herein, and also including removing, inhibiting or otherwise interfering with a gene product from any cellular gene found by the present method to be necessary for viral growth yet nonessential to the cell. For example, DNA viruses or RNA viruses can be targeted. One can readily determine whether cells infected with a selected virus can be selectively removed from a culture through removal of serum by starving cells permissive to the virus of serum (or inhibiting the serum survival factor), adding the selected virus to the cells, adding serum to the culture, and observing whether infected cells die (*i.e.*, by titering levels of virus in the surviving cells with an antibody specific for the virus).

A culture of any animal cell (*i.e.*, any cell that is typically grown and maintained in culture in serum) that can be maintained for a period of time in the absence of serum, can be purified from viral infection utilizing the present method. For example, primary cultures as well as established cultures and cell lines can be used. Furthermore, cultures of
5 cells from any animal and any tissue or cell type within that animal that can be cultured and that can be maintained for a period of time in the absence of serum can be used. For example, cultures of cells from tissues typically infected, and particularly persistently infected, by an infectious virus could be used.

As used in the claims "in the absence of serum" means at a level at which
10 persistently virally infected cells do not survive. Typically, the threshold level is about 1% serum in the media. Therefore, about 1% serum or less can be used, such as about 1%, 0.75%, 0.50%, 0.25%, 0.1% or no serum can be used.

As used herein, "selectively eliminating" cells persistently infected with a virus means that substantially all of the cells persistently infected with the virus are killed such
15 that the presence of virally infected cells cannot be detected in the culture immediately after the elimination procedure has been performed. Furthermore, "selectively eliminating" includes that cells not infected with the virus are generally not killed by the method. Some surviving cells may still produce virus but at a lower level, and some may be defective in pathways that lead to death by the virus. Typically, for cells persistently
20 infected with virus to be substantially all killed, more than about 90% of the cells, and more preferably more than about 95%, 98%, 99%, or 99.99% of virus-containing cells in the culture are killed.

The present method also provides a nucleic acid comprising the regulatory region of any of the genes. Such regulatory regions can be isolated from the genomic sequences
25 isolated and sequenced as described above and identified by any characteristics observed that are characteristic for regulatory regions of the species and by their relation to the start codon for the coding region of the gene. The present invention also provides a construct comprising the regulatory region functionally linked to a reporter gene. Such constructs are made by routine subcloning methods, and many vectors are available into which
30 regulatory regions can be subcloned upstream of a marker gene. Marker genes can be chosen for ease of detection of marker gene product.

The present method therefore also provides a method of screening a compound for treating a viral infection, comprising administering the compound to a cell containing any

of the above-described constructs, comprising a regulatory region of one of the genes comprising any of the nucleotide sequences set forth in SEQ LIST 2, or any homologs thereof, whose inhibition or reduction in expression causes inhibition of viral replication wherein the region is functionally linked to a reporter gene, and detecting the level of the reporter gene product produced, a decrease or elimination of the reporter gene product indicating a compound for treating the viral infection. Compounds detected by this method would inhibit transcription of the gene from which the regulatory region was isolated, and thus, in treating a subject, would inhibit the production of the gene product produced by the gene, and thus treat the viral infection.

Some genes when disrupted by the present method of retrovirus insertion, resulted in over expression of the gene product, and this overexpression inhibited viral replication. Thus the present invention provides a method of screening a compound for effectiveness in treating a viral infection, comprising administering the compound to a cell containing a cellular gene functionally encoding a gene product whose overexpression inhibits reproduction of the virus but does not prevent survival of the cell and detecting the level of the gene product produced, an increase in the gene product indicating a compound effective for treating the viral infection. Typically, an increase will be a 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 125%, 150%, 175%, 200%, 300%, 400%, 500% or higher increase over gene product produced when the compound is not present.

The present invention additionally provides a method of reducing or inhibiting a viral infection in a subject, comprising administering to the subject an amount of a composition that inhibits expression or functioning of a gene product encoded by a gene comprising the nucleic acid set forth in any of SEQ LIST 2, or a homolog thereof, thereby treating the viral infection. Reducing or inhibiting viral infection naturally can include both the initial infection of the subject and the infection of uninfected cells within an already infected subject, e.g. inhibiting viral replication in cells of the subject. The composition can comprise, for example, an antibody that binds a protein encoded by the gene. The composition can also comprise an antibody that binds a receptor for a protein encoded by the gene. Such an antibody can be raised against the selected protein by standard methods as set forth above, and can be either polyclonal or monoclonal, though monoclonal is preferred. Alternatively, the composition can comprise an antisense RNA that binds an RNA encoded by the gene, as described above. Examples of antisense RNA useful therapeutically include the fragments of the nucleic acids described above.

Furthermore, the composition can comprise a nucleic acid functionally encoding an antisense RNA that binds an RNA encoded by the gene. Other useful compositions will be readily apparent to the skilled artisan.

The present invention also provides a method of treating a viral infection in a subject comprising administering to the subject a treatment effective amount of a composition that increases expression of a gene whose over expression reduces or inhibits viral replication. Typically, an increase will be a 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 125%, 150%, 175%, 200%, 300%, 400%, 500% or higher increase over gene product produced when the composition is not present.

10 The present invention further provides a method of reducing or inhibiting a viral infection in a subject comprising mutating *ex vivo* in a selected cell, for example from the subject or from an allogenic source, an endogenous gene comprising a nucleic acid set forth in SEQ LIST 2 whose inhibition or reduction in expression causes inhibition of viral replication, or a homolog thereof, to a gene form incapable of producing a functional gene product of the gene or a gene form producing a reduced amount of a functional gene product of the gene, and placing (or replacing, in the case of the subject's own cells) the cell in the subject, thereby reducing viral infection of cells in the subject. The cell can be selected according to the typical target cell of the specific virus whose infection is to be reduced, prevented or inhibited. A preferred cell for several viruses is a hematopoietic cell. When the selected cell is a hematopoietic cell, viruses which can be reduced or inhibited from infection can include, for example, HIV, including HIV-1 and HIV-2. However, many other virus-cell combinations will be apparent to the skilled artisan.

25 The invention also includes a method of reducing or inhibiting viral infection in a subject comprising mutating *ex vivo* in a selected cell, for example from a subject or an allogenic source, an endogenous gene comprising a nucleic acid set forth in SEQ LIST 2 whose overexpression causes inhibition of viral replication, or a homolog thereof, to a gene form that expresses the gene at a higher level than the endogenous gene, and placing or replacing the cell in the subject. Typically, a higher level can be 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 125%, 150%, 175%, 200%, 300%, 400%, 500% or higher than the non-mutated, endogenous gene. The cell can be selected according to the typical target cell of the specific virus whose infection is to be reduced, prevented or inhibited. A preferred cell for several viruses is a hematopoietic cell. When the selected cell is a hematopoietic cell, viruses which can be reduced or inhibited from infection can

include, for example, HIV, including HIV-1 and HIV-2. However, many other virus-cell combinations will be apparent to the skilled artisan.

The present invention additionally provides a method of increasing viral infection resistance in a subject comprising mutating *ex vivo* in a selected cell, for example from the subject or from an allogenic source, an endogenous gene comprising a nucleic acid set forth in SEQ LIST 2, whose inhibition or reduction in expression increases viral infection resistance, said endogenous gene being mutated to a mutated gene form incapable of producing a functional gene product of the gene or a gene form producing a reduced amount of a functional gene product of the gene, and placing the cell in the subject, thereby increasing viral infection resistance of cells in the subject. The virus can be HIV, particularly when the cell is a hematopoietic cell. However, many other virus-cell combinations will be apparent to the skilled artisan.

Furthermore, the present invention provides a method for isolation of cellular genes utilized in tumor progression. The present invention provides a method of identifying a cellular gene that can suppress a malignant phenotype in a cell, comprising (a) transferring into a cell culture incapable of growing well in soft agar or Matrigel a vector encoding a selective marker gene lacking a functional promoter, (b) selecting cells expressing the marker gene, and (c) isolating from selected cells which are capable of growing in soft agar or Matrigel a cellular gene within which the marker gene is inserted, thereby identifying a gene that can suppress a malignant phenotype in a cell. This method can be performed using any selected non-transformed cell line, of which many are known in the art.

The present invention additionally provides a method of identifying a cellular gene that can suppress a malignant phenotype in a cell, comprising (a) transferring into a cell culture of non-transformed cells a vector encoding a selective marker gene lacking a functional promoter, (b) selecting cells expressing the marker gene, and (c) isolating from selected and transformed cells a cellular gene within which the marker gene is inserted, thereby identifying a gene that can suppress a malignant phenotype in a cell. A non-transformed phenotype can be determined by any of several standard methods in the art, such as the exemplified inability to grow in soft agar, or inability to grow in Matrigel.

The present invention further provides a method of screening for a compound for suppressing a malignant phenotype in a cell comprising administering the compound to a cell containing a cellular gene functionally encoding a gene product involved in

establishment of a malignant phenotype in the cell and detecting the level of the gene product produced, a decrease, inhibition or elimination of the gene product indicating a compound effective for suppressing the malignant phenotype. Detection of the level, or amount, of gene product produced can be measured, directly or indirectly, by any of
5 several methods standard in the art (*e.g.*, protein gel, antibody-based assay, detecting labeled RNA) for assaying protein levels or amounts, and selected based upon the specific gene product.

The present invention also provides a method of screening for a compound for suppressing a malignant phenotype in a cell comprising administering the compound to a
10 cell containing a cellular gene functionally encoding a gene product whose overexpression is involved in suppressing a malignant phenotype in the cell and detecting the level of the gene product produced, an increase in the gene product indicating a compound effective for suppressing the malignant phenotype.

The present invention further provides a method of suppressing a malignant
15 phenotype in a cell in a subject, comprising administering to the subject an amount of a composition that inhibits expression or functioning of a gene product encoded by a gene comprising the nucleic acid set forth in SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:36 or SEQ ID NO:94, or a homolog thereof, or any gene whose overexpression is found by the present method to be
20 involved in suppressing a malignant phenotype in the cell (*e.g.*, any clone designated herein with an "x") thereby suppressing a malignant phenotype. The composition can, for example, comprise an antibody that binds a protein encoded by the gene. The composition can, as another example, comprise an antibody that binds a receptor for a protein encoded by the gene. The composition can comprise an antisense RNA that binds an RNA
25 encoded by the gene. Further, the composition can comprise a nucleic acid functionally encoding an antisense RNA that binds an RNA encoded by the gene.

The present invention further provides a method of suppressing a malignant phenotype in a cell in a subject, comprising administering to the subject an amount of a composition that increases expression of a gene product whose overexpression is involved
30 in suppressing a malignant phenotype in the cell. The gene product can be the product of a gene wherein disruption of an upstream gene by the present vector resulted in overexpression of the downstream gene, and the overexpression of the downstream gene

demonstrated a transformed phenotype. The composition can be, for example, an inhibitor, such as a small molecule inhibitor, of the COX 2 enzyme.

Diagnostic or therapeutic agents of the present invention can be administered to a subject or an animal model by any of many standard means for administering therapeutics or diagnostics to that selected site or standard for administering that type of functional entity. For example, an agent can be administered orally, parenterally (e.g., intravenously), by intramuscular injection, by intraperitoneal injection, topically, transdermally, or the like. Agents can be administered, e.g., as a complex with cationic liposomes, or encapsulated in anionic liposomes. Compositions can include various amounts of the selected agent in combination with a pharmaceutically acceptable carrier and, in addition, if desired, may include other medicinal agents, pharmaceutical agents, carriers, adjuvants, diluents, etc. Parental administration, if used, is generally characterized by injection. Injectables can be prepared in conventional forms, either as liquid solutions or suspensions, solid forms suitable for solution or suspension in liquid prior to injection, or as emulsions. Depending upon the mode of administration, the agent can be optimized to avoid degradation in the subject, such as by encapsulation, etc.

Dosages will depend upon the mode of administration, the disease or condition to be treated, and the individual subject's condition, but will be that dosage typical for and used in administration of antiviral or anticancer agents. Dosages will also depend upon the composition being administered, e.g., a protein or a nucleic acid. Such dosages are known in the art. Furthermore, the dosage can be adjusted according to the typical dosage for the specific disease or condition to be treated. Furthermore, viral titers in culture cells of the target cell type can be used to optimize the dosage for the target cells *in vivo*, and transformation from varying dosages achieved in culture cells of the same type as the target cell type can be monitored. Often a single dose can be sufficient; however, the dose can be repeated if desirable. The dosage should not be so large as to cause adverse side effects. Generally, the dosage will vary with the age, condition, sex and extent of the disease in the patient and can be determined by one of skill in the art. The dosage can also be adjusted by the individual physician in the event of any complication.

For administration to a cell in a subject, the composition, once in the subject, will of course adjust to the subject's body temperature. For *ex vivo* administration, the composition can be administered by any standard methods that would maintain viability of the cells, such as by adding it to culture medium (appropriate for the target cells) and

adding this medium directly to the cells. As is known in the art, any medium used in this method can be aqueous and non-toxic so as not to render the cells non-viable. In addition, it can contain standard nutrients for maintaining viability of cells, if desired. For *in vivo* administration, the complex can be added to, for example, a blood sample or a tissue sample from the patient, or to a pharmaceutically acceptable carrier, e.g., saline and buffered saline, and administered by any of several means known in the art. Examples of administration include parenteral administration, e.g., by intravenous injection including regional perfusion through a blood vessel supplying the tissues(s) or organ(s) having the target cell(s), or by inhalation of an aerosol, subcutaneous or intramuscular injection, topical administration such as to skin wounds and lesions, direct transfection into, e.g., bone marrow cells prepared for transplantation and subsequent transplantation into the subject, and direct transfection into an organ that is subsequently transplanted into the subject. Further administration methods include oral administration, particularly when the composition is encapsulated, or rectal administration, particularly when the composition is in suppository form. A pharmaceutically acceptable carrier includes any material that is not biologically or otherwise undesirable, i.e., the material may be administered to an individual along with the selected complex without causing any undesirable biological effects or interacting in a deleterious manner with any of the other components of the pharmaceutical composition in which it is contained.

Specifically, if a particular cell type *in vivo* is to be targeted, for example, by regional perfusion of an organ or tumor, cells from the target tissue can be biopsied and optimal dosages for import of the complex into that tissue can be determined *in vitro*, as described herein and as known in the art, to optimize the *in vivo* dosage, including concentration and time length. Alternatively, cultured cells of the same cell type can also be used to optimize the dosage for the target cells *in vivo*.

For either *ex vivo* or *in vivo* use, the complex can be administered at any effective concentration. An effective concentration is that amount that results in reduction, inhibition or prevention of the viral infection or in reduction or inhibition of the transformed phenotype of the cells.

A nucleic acid can be administered in any of several means, which can be selected according to the vector utilized, the organ or tissue, if any, to be targeted, and the characteristics of the subject. The nucleic acids, if desired in a pharmaceutically acceptable carrier such as physiological saline, can be administered systemically, such as

intravenously, intraarterially, orally, parenterally, subcutaneously. The nucleic acids can also be administered by direct injection into an organ or by injection into the blood vessel supplying a target tissue. For an infection of cells of the lungs or trachea, it can be administered intratracheally. The nucleic acids can additionally be administered topically, 5 transdermally, etc.

The nucleic acid or protein can be administered in a composition. For example, the composition can comprise other medicinal agents, pharmaceutical agents, carriers, adjuvants, diluents, etc. Furthermore, the composition can comprise, in addition to the vector, lipids such as liposomes, such as cationic liposomes (e.g., DOTMA, DOPE, 10 DC-cholesterol) or anionic liposomes. Liposomes can further comprise proteins to facilitate targeting a particular cell, if desired. Administration of a composition comprising a vector and a cationic liposome can be administered to the blood afferent to a target organ or inhaled into the respiratory tract to target cells of the respiratory tract. Regarding liposomes, see, e.g., Brigham et al. *Am. J. Resp. Cell. Mol. Biol.* 1:95-100 (1989); Felgner et al. *Proc. Natl. Acad. Sci USA* 84:7413-7417 (1987); U.S. Pat. 15 No.4,897,355.

For a viral vector comprising a nucleic acid, the composition can comprise a pharmaceutically acceptable carrier such as phosphate buffered saline or saline. The viral vector can be selected according to the target cell, as known in the art. For example, 20 adenoviral vectors, in particular replication-deficient adenoviral vectors, can be utilized to target any of a number of cells, because of its broad host range. Many other viral vectors are available, and their target cells are known.

Throughout this application, various publications are referenced. The disclosures 25 of these publications in their entireties are hereby incorporated by reference into this application in order to more fully describe the state of the art to which this invention pertains.

EXAMPLES

30

Selective elimination of virally infected cells from a cell culture

Rat intestinal cell line-1 cells (RIE-1 cells) were standardly grown in Dulbecco's modified Eagle's medium, high glucose, supplemented with 10% fetal bovine serum. To

begin the experiment, cells persistently infected with reovirus were grown to near confluence, then serum was removed from the growth medium by removing the medium, washing the cells in PBS, and returning to the flask medium not supplemented with serum. Typically, the serum content was reduced to 1% or less. The cells are starved for serum
5 for several days, or as long as about a month, to bring them to quiescence or growth arrest. Media containing 10% serum is then added to the quiescent cells to stimulate growth of the cells. Surviving cells are found to not be persistently infected cells by immunohistochemical techniques used to establish whether cells contain any infectious virus (sensitivity to 1 infectious virus per ml of homogenized cells).

10

Cellular Genomic DNA Isolation

Gene Trap Libraries: The libraries are generated by infecting the RIE-1 cells with a retrovirus vector (U3 gene-trap) at a ratio of less than one retrovirus for every ten cells. When a U3 gene trap retrovirus integrates within an actively transcribed gene, the
15 neomycin resistance gene that the U3 gene trap retrovirus encodes is also transcribed, thus conferring resistance to the cell to the antibiotic neomycin. Cells with gene trap events are able to survive exposure to neomycin while cells without a gene trap event die. The various cells that survive neomycin selection are then propagated as a library of gene trap events. Such libraries can be generated with any retrovirus vector that has the properties
20 of expressing a reporter gene from a transcriptionally active cellular promoter that tags the gene for later identification.

Reovirus selection: Reovirus infection is typically lethal to RIE-1 cells but can result in the development of persistently infected cells. These cells continue to grow while producing infective reovirus particles. For the identification of gene trap events that
25 confer reovirus resistance to cells, the persistently infected cells must be eliminated or they will be scored as false positives. We have found that RIE-1 cells persistently infected with reovirus are very poorly tolerant to serum starvation, passaging and plating at low density. Thus, we have developed protocols for the screening of the RIE-1 gene trap libraries that select against both reovirus sensitive cells and cells that are persistently infected with
30 reovirus.

1. RIE-1 library cells are grown to near confluence and then the serum is removed from the media. The cells are starved for serum for several days to bring them to quiescent or growth arrest.

2. The library cells are infected with reovirus at a titer of greater than ten reovirus per cell and the serum starvation is continued for several more days.
3. The infected cells are passaged, (a process in which they are exposed to serum for three to six hours) and then starved for serum for several more days.
- 5 4. The surviving cells are then allowed to grow in the presence of serum until visible colonies develop at which point they are cloned by limiting dilution.

MEDIA: DULBECCO'S MODIFIED EAGLE'S MEDIUM, HIGH GLUCOSE (DME/HIGH) Hyclone Laboratories cat. no. SH30003.02.

NEOMYCIN: The antibiotic used to select against the cells that did not have a U3 gene trap retrovirus, e.g. GENETICIN, from Sigma. [cat. no. G9516].

RAT INTESTINAL CELL LINE-1 CELLS (RIE-1 CELLS): These cells are from the laboratory of Dr. Ray Dubois (VAMC). They are typically cultured in Dulbecco's Modified Eagle's Medium supplemented with 10% fetal calf serum.

REOVIRUS: Laboratory strains of either serotype 1 or serotype 3 are used. They were originally obtained from the laboratories of Bernard N. Fields (deceased). These viruses have been described in detail.

RETROVIRUS: The U3 gene trap retrovirus used here were developed by Dr. Earl Ruley (VAMC) and the libraries were produced using a general protocol suggested by him.

SERUM: FETAL BOVINE SERUM Hyclone Laboratories cat. no. A-1115-L.

20

Identification Tags for Isolated Nucleic Acids

Genomic sequences, tagged with a vector, such as the U3 gene trap vector, are given a number corresponding to the genomic library of mutant cells from which the sequence was isolated., and a letter indicating a unique member of the library. More than one sequence with the same number and letter indicates multiple, unique sequences obtained from the genome surrounding the vector insert that "tagged" the gene. Such genomic sequences are obtained using vector-based primers, from which sequencing occurs 3' to 5' or 5' to 3'. In the former case, to recover the orientation of the gene into which the vector inserted, the sequence derived from the vector primer must be reversed and complemented. Such reverse complement sequences are designated "rE". In the case of genome sequencing from a primer that occurs 5' to 3' (i.e. the primer is at the 3' end of the vector), no changes are needed, since the derived sequence is the sequence as it appears

in the gene disrupted. Such sequences are designated "B4". Homologies indicated below each genomic sequence are in the positive direction, unless explicitly noted to be on the negative strand. As an example, SEQ ID NO. 27 comprises a nucleic acid sequence encoding a novel polypeptide on the positive strand, while the negative strand encodes
 5 ferritin.

	SEQ ID NO:	Lab Designation
	1	32-3-2#1E/-rE
	2	L191B2E#1-RE
10	3	L191B2E#3+-rE
	4	21-5-9E-RE
	homology to: emb/AL021154/HS15005 human DNA sequence	
	5	14A14E-rE
15	6	4cx-b4
	7	5a-b4
	8	6BSA12-B4
	9	X7B/B4
	10	x27b4f_1
20	11	12C#A-rE
	12	10-3b(5/2/96)/-rE
	13	10_4B_4-rE
	14	6BE60-rE
25	homology to: alpha-trophomyosin	
	15	19D3E-rE
	16	L19D16E-rE
	17	2b_rE
	18	14_24_#6-rE
30	19	7A7'-rE
	homology to: annexin II/dynein I	
	20	L12cx#6-rE
	homology to: gb:X51760 human zinc finger protein ZFP-36	
35	21	L12cx#11-rE

- 22 19D5E-rE
homology to: 6-pyruvoyl-tetrahydropterin synthase (gb/M77850/RAT6PTHS)
- 23 12_3b#7-rE
- 5 24 12_3B#8-RE
homology to: gb/AA871174/vq32a08.r1 Barskad bowel MPLRBg Mus musculus cDNA clone 10959265'
- 25 9B27-2-E
homology to: RAT LOCUS RNU53922 04-MAY-1996; Rattus norvegicus DnaJ-like protein (RDJ1) mRNA, complete Cds, ACCESSION U53922 (on negative strand)
- 10 26 x15-rE
- 27 X11-rE
homology to: ferritin H (on the negative strand)
- 15 28 X20-rE
homology to: LOCUS RATGL5A Rat NICER element (GL5-14)5' long terminal repeat, Acc.No. M59028 M33535N1D
- 20 29 X4-rE
- 30 14A7E-rE
homology to: MMSMAD7 3681 bp mRNA ROD 31-JUL-1998 DEFINITION Mus musculus mRNA for Mad-related protein Smad7 ,149 bases
- 25 31 14A13E-rE
- 32 14_7#2E-rE
homology to: N-acetylglucosaminyltransferase I
- 33 12CX#6-rE
- 30 homology to: gb|AA522204|AA522204 vf98g09_r1 Soares mouse mammary gland NbMMG Mus musculus cDNA clone 851872; also 5' similar to gb X51760 zinc finger protein ZFP-36 (HUMAN), gb L20450 Mus musculus DNA-binding protein mRNA, complete cds (MOUSE); Length = 442, 925 bases (shares homology with SEQ ID NO:20)
- 35 34 12C_2B#9E-rE
- 35 12CX#11E-rE
- 36 x5-rE
- 37 8C5_11-rE
- 38 191E2E-rE
- 40 39 19_7AE-rE

- 40 19_9BE-rE
homology to: LOCUS HS347M6 56583 bp DNA PRI 14-JAN-1998 Human DNA
sequence from PAC 347M6 on chromosome Xq22, CSTF2 (Cleavage Stimulation Factor,
CF-1, Polyadenylation Factor) 64 kD subunit gene
- 5 41 191E9E-rE
42 191E8E-rE
43 14C_2E/-rE
homology to: gb/H31084/EST104778 Rattus sp. cDNA - 5' end similar to signal
10 recognition particle subunit(19kDa) (on negative strand)
- 44 14H1E-rE
45 14G3E-rE
46 14G_2E-rE
15 47 6_3_6_2E/-rE
homology to: Rattus norvegicus cis-golgi gp130 (on negative strand); and
a HUMAN EST (on positive strand) AI127398; qb70g11.x1 Soares fetal heart NbHH19W
Homo sapiens cDNA clone (1705508 3' mRNA sequence)
- 20 48 14H4E/-rE
49 18A_8_4E-rE
50 18A_8_1E-rE
51 SCB2_19E-rE
52 L197B3E-rE
25 53 L195C5E-rE
homology to: H. pylori and C.jejuni
- 54 21_5_7E-rE
homology to: id3 gene; emb|AL021154|HS150O5 Human DNA sequence from clone
30 150O5; HTGS phase 1 [Homo sapiens]; containing the E2F2 gene for transcription factor
E2F-2 and the ID3 gene for Inhibitor of DNA binding 3 (dominant negative
helix-loop-helix protein), 1R2, Length = 133667, 971 bases
- 55 L195B1E-rE
35 homology to: vK72b07.s1 Knowles Solter mouse 2 cell Mus musculus cDNA clone
960181 5'
- 56 L194c4E-rE
57 L193A1E#A-rE
40 58 L192A3E-rE
59 L1739E-rE

- 60 L192B3E#13-rE
contains sequence identical to: insulin growth factorII/mannose-6-phosphate receptor
- 61 3 2 4 rE
5 located in the same region of the genome as calcyclin, but the gene is "read" in the
opposite direction
- 62 36 7 1 a-rE
63 36 5 1 4 a-rE
64 34 25 5a-rE
10 rat satellite DNA (RATRSSID 93 bp, ROD 12-MAR-1984)
- 65 34 24-126/rE
homology to:
15 **HSU49928** (3096 bp mRNA) PRI 06-APR-1998, Homo sapiens TAK1 binding protein
(TAB1) mRNA, complete cds, ACCESSION U49928 NID g1401125, and
HS333H23 (142274 bp DNA) HTG 17-JUL-1998 Human DNA sequence
- 66 34 23-1/rE
67 36 5 2-6/rE
20 68 36 5 2-196/rE
69 34 23-3/rE
homology to: gb|L16546|RATAP1X Rat P-glycoprotein (mdr1b) gene
70 34 25 23-rE
71 36 5 2-196/rE
25 72 31 3 9/rE
homology to: AA798638 568 bp mRNA EST 10-FEB-1998, vw34b06.r1 Soares mouse
mammary gland NbMMG Mus musculus cDNA clone1245683 5, mRNA sequence, 824
bases.
73 31 3 6-2-rE
30 74 31 3 17-rE
75 31 3 5-rE
homology to: AF046001 2347 bp mRNA PRI 19-FEB-1998, Homo sapiens zinc finger
transcription factor (ZNF207) mRNA, complete Cds, 833 bases.
- 35 76 31 3 15#1/rE
77 24 3 5#1/rE
78 31 4 4#1/rE
79 31 3 19#2/rE
80 31 4 5#1/rE
40 81 24 9 3#2/rE
82 L24_26_1-BL

homology to: AI045472 396 bp mRNA EST 06-JUL-1998, UI-R-C1-jz-h-09-0-UI.s2
UI-R-C1 Rattus norvegicus cDNA clone UI-R-C1-jz-h-09-0-UI 3', mRNA sequence.

83 L24_26_1-B4
5 84 L22_5A1/rE
85 L24_3_2B/rE
86 L24 4-2/rE
87 L24 5-2/rE
88 L24 5-3/rE
10 89 (15-)L28AP/rE

90 L24 26-10/rE

homology to: LOCUS R06687 403 bp mRNA EST 03-APR-1995; yf10a10.r1 Soares fetal
liver spleen 1NFLS Homo sapiens cDNA clone 126426 5'

15 91 L24 26-2A/rE
92 L24 26-2B/rE

homology to: gb|AA590026|AA590026 vm22g03.r1 Knowles Solter mouse blastocyst B1
Mus musculus cDNA clone 990964 , 459 bases, 139A; and Rattus norvegicus Eker
rat-associated intracisternal-A particle element

20

93 14 7#2E-rE

homology to: N-acetylglucosaminyltransferase I; this sequence shares homology with SEQ
ID NO:32.

25 94 x18
95 31_3_9-rE
96 31_3_6_2-rE
97 31_3_17-rE
98 31_3_15#1-rE
30 99 24_3_5#1-rE
100 31_4_4#1-rE
101 31_3_19#2-rE
102 31_4_5#1-rE
103 24_9_3#2-rE
35 104 14XD#12E-rE
105 70A-rE
106 31-3-4-rE
107 3_6_9-NeoG-rE

- 108 31_4_2-rE
- 109 3_2_13-rE
homology to: calcyclin
- 5 110 3_2_4-E
homology to: pistre-alpha 1 (with homology to calcyclin on negative strand)
- 111 L25-10/-rE
homology to: calcyclin
- 10 112 L24-4-3/-rE
- 113 L24-9-1-rE
rat id sequence
- 15 114 17-L25-27#7-rE
homology to: calcyclin
- 115 L21C1E-rE
homology to: calcyclin
- 20 116 L24-5-3BE-rE.
homology to:
LOCUS H32572 310 bp mRNA EST 08-SEP-1995 EST107805 Rat PC-12 cells, untreated
Rattus sp cDNA 5' end, ACCESSION H32572, and
- 25 LOCUS AA858747 470 bp mRNA EST 10-MAR-1998 UI-R-A0-bb-e-01-0-UI.s1 UI-R-A0
Rattus norvegicus cDNA clone UI-R-A0-bb-e-01-0-UI, 3' similar to gb|AA473081|AA473081
vd44b07-r1 Barstead MPLRB1 Mus musculus cDNA clone 803413 5' mRNA sequence
- 117 L24-4-2BE-rE
- 30 homology to: LOCUS MMU51002 6495 bp DNA ROD 16-JAN-1997 Mus musculus Dlx-2
gene, complete cds, ACCESSION U51002 NID g1477589
- 118 17-3-3B-B4
- 119 L24-26-3/-rE
- 35 homology to: RNU23776, DNA ROD 10-AUG-1995, Rattus norvegicus Eker rat-associated
intracisternal-A particle element
- 120 12_2B#2-rE
- 121 05-17-3-3He-1-T7
- 40 122 21_5_8E-rE
homology to: emb|AL021154|HS150O5 Human DNA sequence from clone 150O5;
1p36_13-36_22, contains the E2F2 gene for transcription factor E2F-2 and the ID3 gene for
Inhibitor of DNA binding 3(dominant negative helix-loop-helix protein, 1R2, Length =
133667, 971 bases
- 45 123 X18H-t7

124	18A_8_4E-rE
125	L24-5-2BE-rE
126	L24-4-2AE-rE
127	L24-10-1BE-rE

5

Genes Necessary for Viral Infection

Some of the isolated sequences disclosed here comprise sequence encoding the following proteins: tristetraprolin (human ZFP-36), 6-pyruvoyltetrahydropterin synthase, a eukaryotic DnaJ-like protein, ID3 (inhibitor of DNA binding 3), N-acetylglucos-
10 aminyltransferase I (mGAT-1), cleavage stimulation factor (CSTF2), TAK1 binding protein, human zinc transcription factor ZPF207, Dlx2, Smad7 (Mad-related protein), and P-glycoprotein (mdr1b).

Isolation of cellular genes that suppress a malignant phenotype

15 We have utilized a gene-trap method of selecting cell lines that have a transformed phenotype (are potentially tumor cells) from a population of cells (RIE-1 parentals) that are not transformed. The parental cell line, RIE-1 cells, does not have the capacity to grow in soft agar or to produce tumors in mice. Following gene-trapping, cells were screened for their capacity to grow in soft agar. These cells were cloned and genomic sequences were obtained
20 5' or 3' of the retrovirus vector, i.e. SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:36 or SEQ ID NO:94; sequences designated with an "x" in the clone name). All of the cell lines behave as if they are tumor cell lines, as they also induce tumors in mice.

Of the cell lines, two are associated with the enhanced expression of the prostaglandin
25 synthetase gene II or COX 2. It has been shown that disruption of gene function by retroviral targeting of an upstream gene has lead to increased expression of a downstream gene product, COX 2. When a small molecule inhibitor of COX 2 enzyme was added, reversion of the transformed phenotype occurred. The COX 2 gene has been found to be increased in pre-malignant adenomas in humans and overexpressed in human colon cancer. Inhibitors of COX
30 2 expression also arrests the growth of the tumor. One of the cell lines, x18 (SEQ ID NO:94), has disrupted a gene that is now represented in the EST (dbest) database, but the gene is not known (not present in GenBank).

Each of the genes from which the provided nucleotide sequences is isolated (and all clones designated with an "x") represents a tumor suppressor gene. The mechanism by which the disrupted genes may suppress a transformed phenotype is at present unknown. However, each one represents a tumor suppressor gene that is potentially unique, as none of the genomic sequences correspond to a known gene. The capacity to select quickly tumor suppressor genes may provide unique targets in the process of treating or preventing (potential for diagnostic testing) cancer.

Isolation of entire genomic genes

10 An isolated nucleic acid of this invention (whose sequence is set forth in any of SEQ ID NO:1 through SEQ ID NO:127), or a smaller fragment thereof, is labeled by a detectable label and utilized as a probe to screen a rat genomic library (lambda phage or yeast artificial chromosome vector library) under high stringency conditions, *i.e.*, high salt and high temperatures to create hybridization and wash temperature 5-20°C. Clones are isolated and
15 sequenced by standard Sanger dideoxynucleotide sequencing methods. Once the entire sequence of the new clone is determined, it is aligned with the probe sequence and its orientation relative to the probe sequence determined. A second and third probe is designed using sequences from either end of the combined genomic sequence, respectively. These probes are used to screen the library, isolate new clones, which are sequenced. These
20 sequences are aligned with the previously obtained sequences and new probes designed corresponding to sequences at either end and the entire process repeated until the entire gene is isolated and mapped. When one end of the sequence cannot isolate any new clone, a new library can be screened. The complete sequence includes regulatory regions at the 5' end and a polyadenylation signal at the 3' end.

25

Isolation of cDNAs

An isolated nucleic acid (whose sequence is set forth in any of SEQ ID NO:1 through SEQ ID NO:127), or a smaller fragment thereof, or additional fragments obtained from the genomic library, that contain open reading frames, is labeled by a detectable label and utilized
30 as a probe to screen a portions of the present fragments, to screen a cDNA library. A rat cDNA library obtains rat cDNA; a human cDNA library obtains a human cDNA. Repeated screens can be utilized as described above to obtain the complete coding sequence of the gene

from several clones if necessary. The isolates can then be sequenced to determine the nucleotide sequence by standard means such as dideoxynucleotide sequencing methods.

Serum survival factor isolation and characterization

5 The lack of tolerance to serum starvation is due to the acquired dependence of the persistently infected cells for a serum factor (survival factor) that is present in serum. The serum survival factor for persistently infected cells has a molecular weight between 50 and 100 kD and resists inactivation in low pH (pH2) and chloroform extraction. It is inactivated by boiling for 5 minutes [once fractionated from whole serum (50 to 100 kD fraction)], and
10 in low ionic strength solution [10 to 50 mM].

The factor was isolated from serum by size fraction using centriprep molecular cut-off filters with excluding sizes of 30 and 100 kd (Millipore and Amnicon), and dialysis tubing with a molecular exclusion of 50 kd. Polyacrylamide gel electrophoresis and silver staining was used to determine that all of the resulting material was between 50 and 100 kd,
15 confirming the validity of the initial isolation. Further purification was performed on using ion exchange chromatography, and heparin sulfate adsorption columns, followed by HPLC. Activity was determined following adjusting the pH of the serum fraction (30 to 100 kd fraction) to different pH conditions using HCl and readjusting the pH to pH 7.4 prior to assessment of biologic activity. Low ionic strength sensitivity was determined by dialyzing
20 the fraction containing activity into low ionic strength solution for various lengths of time and readjusting ionic strength to physiologic conditions prior to determining biologic activity by dialyzing the fraction against the media. The biologic activity was maintained in the aqueous solution following chloroform extraction, indicating the factor is not a lipid. The biologic activity was lost after the 30 to 100 kd fraction was placed in a 100°C water bath for 5
25 minutes.

Isolated nucleic acids

Tagged genomic DNAs isolated were sequenced by standard methods using Sanger dideoxynucleotide sequencing. The sequences were run through computer databanks in a
30 homology search. These genes can be therapy targets particularly because disruption of one or both alleles results in a viable cell.

Seq. List.
insert here
per p# 7/13